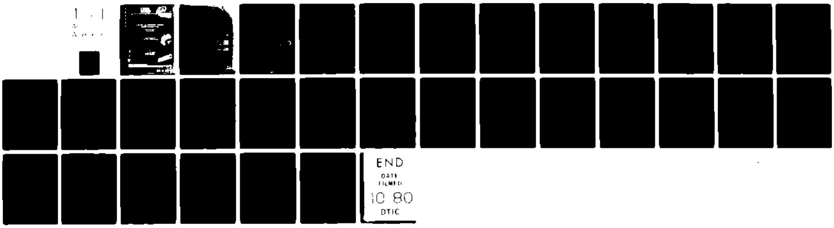
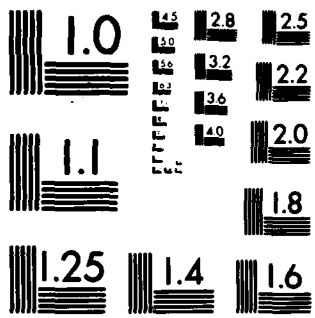
AD-A089 619 DECISION RESEARCH EUGENE OR F/G 5/10
HOW WELL DO PROBABILITY EXPERTS ASSESS PROBABILITIES?(U)
AUG 80 S LICHTENSTEIN, B FISCHHOFF N00014-80-C-0150
UNCLASSIFIED PTR-1092-80-8 NL

1 of 1



END
DATE
FILMED
10 80
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12)

Technical Report PTR-1092-80-8
Contract No. N00014-80-C-0150
Work Unit No. NR 197-064
August 1980

HOW WELL DO PROBABILITY EXPERTS ASSESS PROBABILITIES?

SARAH LICHTENSTEIN
BARUCH FISCHHOFF

DECISION RESEARCH
A BRANCH OF PERCEPTRONICS

Prepared For:

OFFICE OF NAVAL RESEARCH
Department of the Navy
800 North Quincy Street
Arlington, Virginia 22217

PERCEPTRONICS

6271 VARIEL AVENUE • WOODLAND HILLS • CALIFORNIA 91367 • PHONE (213) 884-7470

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
(14) PTR-1092-80-8

2. GOVT ACCESSION NO.
AD-A089 619

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
(6) How well do probability experts assess probabilities?

5. TYPE OF REPORT & PERIOD COVERED
(9) Technical Report, 1/80 - 8/80

6. PERFORMING ORG. REPORT NUMBER
Jan - Aug 80

7. AUTHOR(s)
(10) Sarah Lichtenstein & Baruch Fischhoff

8. CONTRACT OR GRANT NUMBER(s)
(15) N00014-80-C-0150

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Decision Research
A Branch of Perceptronics
1201 Oak Street, Eugene, Oregon 97401

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Work Unit NR197-064

11. CONTROLLING OFFICE NAME AND ADDRESS
Office of Naval Research
Arlington, Virginia 22217

12. REPORT DATE
(11) August 1980

13. NUMBER OF PAGES
22

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
(18) 31

15. SECURITY CLASS. (of this report)
unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

none

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Calibration
probability assessment

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Past research on people's ability to assess probabilities has shown two common errors, overconfidence in one's knowledge and insensitivity to task difficulty. This research has created a new class of experts: those who have studied probability assessors and who are aware of the common errors. The performance of eight such experts is here compared to the performance of twelve untrained subjects and fifteen who had previously received training in probability assessment. All subjects responded to 500 general-knowledge items whose difficulty could be measured *a priori* from the item context. The experts appeared to (over)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

411215

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

cont'd

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

have overcorrected for the overconfidence error: they were notably underconfident, whereas the untrained subjects were overconfident and the trained subjects were mixed. The experts were more sensitive than the other two groups to variations in item difficulty. However, even they showed a substantial insensitivity to difficulty, relative to ideal performance. Introspection suggests that this second error would be hard to overcome.

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

Subjective assessment of probabilities has in recent years increasingly been recognized as an integral part of decision making, both personal (cf. Jungermann, 1980) and public (e.g., the "Rasmussen Report," U.S. Nuclear Regulatory Commission, 1975). This recognition has led to a burgeoning research literature on people's abilities to make such assessments (see Lichtenstein, Fischhoff & Phillips, 1977). Typically, participants in this research are presented with a series of two-alternative, forced-choice questions. For each question, the assessor first chooses the correct alternative and then assesses the probability that the chosen alternative is in fact correct. Analyses of these data have focused on calibration. An assessor is well calibrated if, over the long run, for all alternatives assigned a given probability, the proportion of true alternatives is equal to the probability assigned. Thus, for example, just 70% of all alternatives assigned a probability of .7 should be correct. When the assessed probabilities are larger than the proportions correct (e.g., 90% confident but only 75% correct), the assessors are called "overconfident." The reverse situation is called "underconfidence."

Two robust findings have emerged from this research. First, people are usually overconfident; they believe they know more than they actually know. Such overconfidence has been demonstrated in a variety of tasks (Fischhoff & Slovic, 1980), response modes (Fischhoff, Slovic & Lichtenstein, 1977), and subject populations (Wright, Phillips, Whalley, Choo, Ng, Tan & Wisuda, 1978; Cambridge & Shreckengost, Note 1).

The second general finding is that the degree of overconfidence is related to the overall difficulty of the task. People are most overconfident with the hardest tasks (Clarke, 1960; Nickerson & McGoldrick,

1965; Pitz, 1974). As task difficulty decreases, so does overconfidence, until, with quite easy tasks, people are underconfident (Lichtenstein & Fischhoff, 1977). Apparently, people are insufficiently sensitive to task difficulty, and fail to shift the distribution of their probabilistic responses as much as they should as task difficulty changes.

Recent attempts to reduce overconfidence by training (Lichtenstein & Fischhoff, 1980) or improved task design (Koriat, Lichtenstein & Fischhoff, 1980) have been moderately successful. However, no one has managed to enhance sensitivity to task difficulty.

All this research has produced a new kind of expertise: people who have studied probability assessors and who are aware of common errors. The present study explores such experts' ability to use their knowledge to overcome the errors exhibited by naive assessors. Their performance is here compared with that of naive subjects and of subjects who had previously been trained to be well calibrated.

Method

Subjects

Experts. The eight expert subjects, five males and three females, included the present authors, two of their research assistants, and four other psychologists who have done research in probability assessment.[1] All reported having read the research literature on calibration and overconfidence.

Trained subjects. In a previous paper (Lichtenstein & Fischhoff, 1980), we reported the results of two studies in which we trained 24 subjects to be well calibrated, using individualized feedback about calibration after each of 3 or 11 sessions of 200-item, general-knowledge tests. Sixteen (8 females and 8 males) of those 24 subjects agreed to

serve in the present experiment. Of these, six had received 11 sessions of previous training, during which they had responded to the 500 items used in the present study, randomly intermixed with 2,500 items covering other topics. However, since subjects had not been told the correct answer to any item during their training, and since a year had elapsed between the end of training and the present experiment, we felt that having previously seen the test items would not significantly affect their present performance. The other ten trained subjects had received only three sessions of training, and thus had been exposed to only a small fraction of the items used here. Three to six months had elapsed between their training and the present experiment.

Untrained subjects. The untrained subjects were 13 people (9 males and 4 females) who responded to a job listing at the University of Oregon branch of the State Employment Division.

Stimuli

The 500 items were of three types. The first 289 items listed pairs of continents, countries, states, or cities; the task was to indicate which was more populous (e.g., [a] Las Vegas, [b] Miami; [a] Helsinki, Finland, [b] Milan, Italy). The next 111 items listed a base city followed by two other cities; the task was to indicate which of the two alternative cities was farther in distance from the base city (e.g., Melbourne: [a] Rome, [b] Tokyo). The final 100 items listed two historical events; the task was to indicate which event happened first (e.g., [a] Magna Carta signed, [b] Mohammed born).

The items were selected by our secretaries from almanacs, under general (and vague) instructions not to make the test *too* hard or *too* easy and to avoid deceptive items (i.e., those that might be answered incorrectly by most people).

Difficulty. In previous research on the relationship between task difficulty and calibration, difficulty was defined either intuitively (based on subjects' presumed familiarity with the topics; Pitz, 1974), or on the basis of subjects' performance in the experiment (i.e., items for which more subjects chose the correct alternative were taken as easier; Clarke, 1960; Lichtenstein & Fischhoff, 1977). The latter strategy leads to an artifactual inflation of the difference in calibration between easy and hard tests that is difficult to separate from valid effects.

To avoid this artifact, the present experiment was designed to define item difficulty *a priori*. Each item involved two numbers: two populations, two distances, or two years. We assumed that the more similar these two numbers, the harder the item is likely to be. To get a measure of difficulty, we formed the ratio of the larger number to the smaller number (for the historical events, the ratios were formed from the number of years elapsed since the events occurred). The 250 items with the largest ratios were designated as *easy*; the rest were called *hard*. The ratios varied from 1.01 to 78.79; the median, at which the hard/easy division was made, was 1.84.

Instructions. The instructions were brief:

> For each question select the answer you believe to be correct (your best guess). Then assess the probability that your answer is, in fact, correct. This probability can be any number from .50 to 1.0. It can be interpreted as your degree of certainty about the correctness of your answer. For example, if you respond that the probability is .6, it means that you believe that there are about 6 chances out of 10 that your answer is correct.

> A response of 1.0 means that you are absolutely certain that your answer is correct. A response of .5 means that your answer is as likely to be right as wrong. Since there are two possible answers, if you make a pure guess, you would probably be right about 1/2 of the time--thus, .50 would be an appropriate probability for such guesses. Write your probability in the space provided on the answer sheet.
>
> To repeat, your probability is a measure of your degree of certainty that your chosen alternative is the correct alternative. It is a number from .50 to 1.0, where .50 means complete uncertainty and 1.0 means complete certainty.

In addition, the expert and trained subjects (all of whom knew about calibration) were asked to be as well calibrated as possible.

All subjects were run individually. The trained and expert subjects who did not live in Eugene were contacted by mail. The trained and untrained subjects were paid for their participation; the experts were not.

Results

Two subjects, one untrained and one trained, apparently misunderstood the instructions for the middle section of the test; instead of picking the city <u>farthest</u> from the base city, they seemed to have picked the <u>closest</u> city. For those 111 items, the untrained subject selected the correct alternative only 36% of the time and the trained subject only 8% of the time. Furthermore, they erred most often on the easiest items. These subjects were dropped from the study. The results that follow are thus based on 8 expert, 15 trained, and 12 untrained subjects.

Performance measures

Table 1 shows the mean and range for several measures of subjects' performance. The experts were most knowledgeable, correctly identifying an average of 75% of the right answers, but they were not the most confident.

Overconfidence. A measure of overall overconfidence is the signed difference between the mean assessed probability and the proportion of correct alternatives chosen. A positive sign indicates overconfidence; a negative sign, underconfidence. By this measure, the untrained group was predominantly overconfident; only one of the 12 subjects was underconfident. Similar overconfidence has been found in previous studies with untrained subjects (Lichtenstein et al., 1977). In contrast, the experts were all underconfident. The trained group was more varied: five were overconfident and ten were underconfident. Figure 1 shows the calibration for the three groups; the overconfidence of the untrained group (represented by a curve falling below the diagonal) and the underconfidence of the experts are readily apparent.

Calibration. A measure of calibration proposed by Murphy (1972, 1973) and used in our previous work (Lichtenstein & Fischhoff, 1977; in press) is the mean squared difference between the assessed probabilities and the corresponding proportions of correct answers, weighted by the number of responses at each point. It measures the mean squared vertical distance, in a plot like Figure 1, between the points and the diagonal.[2] The calibration scores appear in the last row of Table 1. The expert and untrained groups were indistinguishable ($t = .61$; $p > .5$), whereas the trained group was significantly better than both the others ($t = 2.44$; $p = .02$).

Table 1

Means (and Ranges) of Performance Measures

for all 500 Items

Measure	Group		
	Untrained	Trained	Experts
Proportion Correct	.684 (.630 to .837)	.667 (.574 to .786)	.750 (.676 to .806)
Mean Response	.741 (.664 to .837)	.648 (.559 to .764)	.682 (.629 to .729)
Overconfidence	+.057 (-.008 to +.165)	-.029 (-.121 to +.046)	-.068 (-.123 to -.004)
Calibration	.0118 (.0013 to .0339)	.0055 (.0009 to .0176)	.0100 (.0008 to .0172)

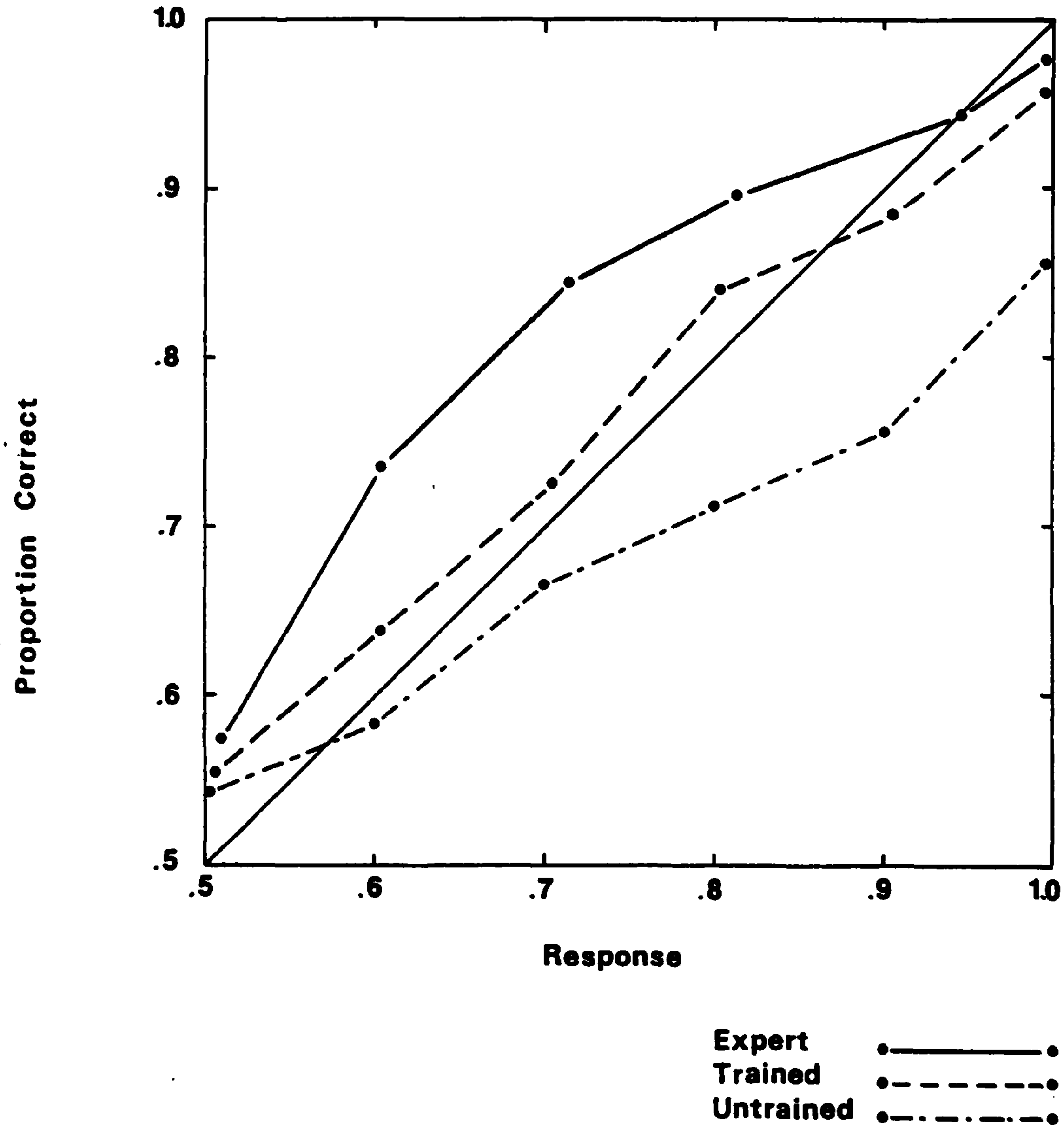


Figure 1. Calibration curves for the three groups of subjects. The responses of all subjects within each group were combined. In addition, all responses less than 1.0 were grouped into categories: .5 to .59, .6 to .69, ... , .9 to .99. The proportion correct in each category is here plotted against the mean response in each category.

<u>Use of 1.0</u>. The probabilistic response of 1.0 indicates complete certainty, and previous research (Fischhoff, Slovic & Lichtenstein, 1977) has shown that people use it too often when they are, in fact, wrong. The untrained subjects replicated this finding. They used 1.0 for 22.4% of their responses, of which only 85% were correct (Figure 1). The trained and expert groups were markedly superior, using 1.0 for 9.6% and 7.7% of their responses, respectively, and getting 96% and 97% of those correct. Two experts never used 1.0; one used it only for correct alternatives.

<u>Difficulty</u>

Our procedure for separating items into "easy" and "hard" tests proved quite successful. Over all 35 subjects, the percent correct for the easy items was 81.4; for the hard items it was 57.8. For all subjects but one, the percent correct on the easy items exceeded the percent correct on the hard items by at least 16 percentage points.

The effect of difficulty on calibration was striking for all subjects. On the hard items, subjects were notably overconfident (or at least much less underconfident; six of the eight experts were still slightly underconfident on the hard items). On the easy items, even 10 out of the 12 untrained subjects were underconfident. The group calibration curves for the hard and easy items are shown in Figure 2.

<u>Use of .5</u>. A response of .5 should represent a "pure guess," as likely to be wrong as right. But for easy items the percentage correct when our subjects responded .5 was substantially greater than 50% for all three groups (experts, 59.3%; trained subjects, 58.5%; and untrained subjects, 60.0%). Some experts' use of the response .5 on the easy items was

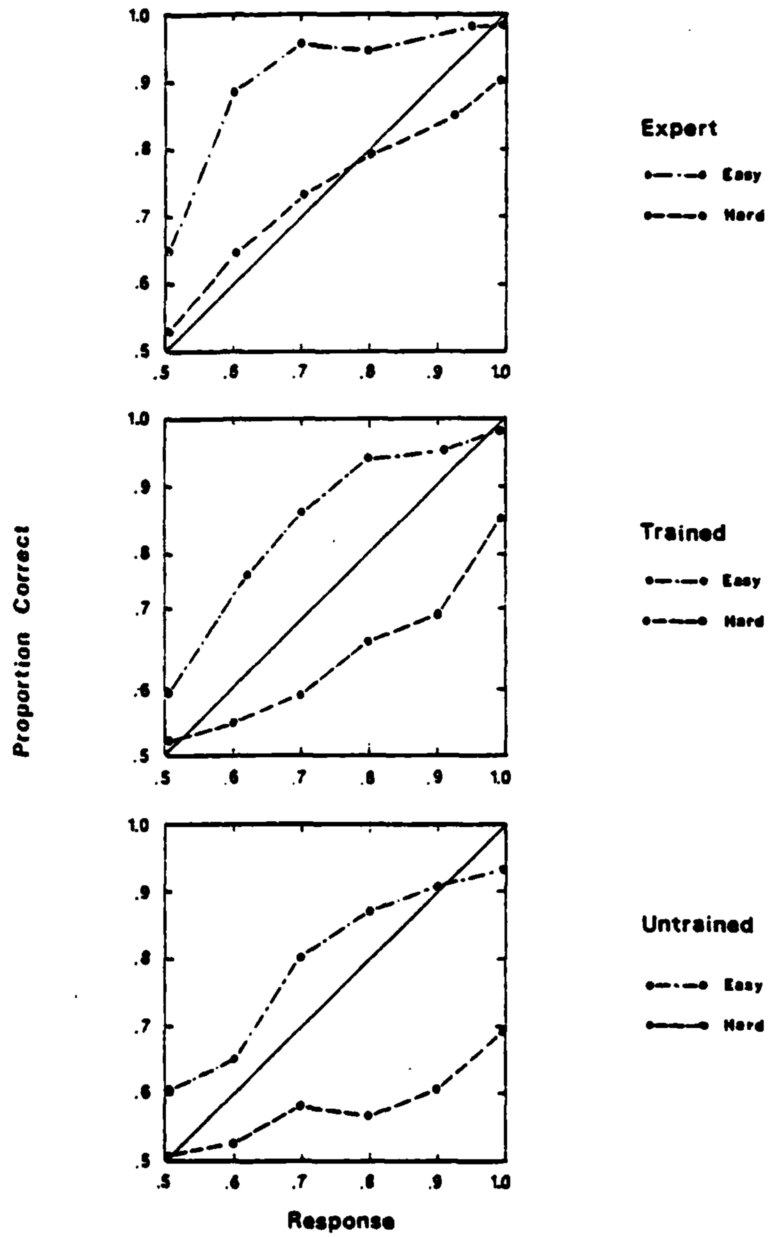


FIGURE 2. Calibration curves for the easy and hard tests, across all subjects within each group.

particularly discouraging: five of the experts responded .5 for 19% of the easy items and got 65% of them right, well above chance performance. The other three experts used artificial strategies for their .5 responses. One did not select an alternative when responding .5 (for data analyses, the computer alternated between a and b). Another always chose alternative a when responding .5, and the third adopted a constant-response strategy near the beginning of the task. Two of the trained group also made strategic choices of alternatives when responding .5. The others got 61% of their easy .5's right. It appears that even those well-schooled in the meaning of ".5" tend to choose the correct answer to easy questions they think they don't know.

Difficulty, overconfidence, and calibration. Consistent with previous research, the subjects in this study tended to be overconfident with hard items and underconfident with easy items. A necessary (but not sufficient) condition for good calibration is that the assessor be neither over- nor underconfident. The strong relationship between difficulty and overconfidence suggests that there is an "ideal" difficulty level for which an assessor will be neither over- nor underconfident and thus will be best calibrated. In Figure 3, overconfidence is plotted against percentage correct for each subject on the hard and easy tests. The straight line in each plot connects the group means. These data suggest that the untrained group might be best calibrated on a test on which they would get about 78% of the items correct. Lichtenstein and Fischhoff (1977) estimated this cross-over point at "approximately 80%" (p. 179) for a different group of untrained subjects. The trained subjects might do best on a test with 63% correct (close to the 68% they achieved in this test and the 67% they scored on the last round of their previous training),

and the experts would require an even more difficult test, 58% correct.

This reasoning suggests that, despite the apparently clear-cut results shown in Figure 1, we cannot unequivocally characterize an assessor or group as "better calibrated" than another without taking into account the relationship between difficulty and overconfidence. Perhaps the trained group weren't better; they were just lucky to receive a test with an overall difficulty level about the same as the difficulty of their training and thus close to their ideal.

Sensitivity to changes in difficulty. Assessors would perform better if they were more sensitive to item difficulty. Ideally, their response distributions would shift enough to make changes in mean response equal to changes in proportion correct. Completely insensitive assessors would maintain the same response distribution for all difficulty levels. Letting Ψ stand for mean assessed probability and P stand for proportion correct, and using subscripts h and e for the hard and easy tests, the ratio:

$$\frac{\Psi_e - \Psi_h}{P_e - P_h}$$

expresses the degree of sensitivity a subject shows to changes in difficulty. For ideal sensitivity, the index would be equal to 1.0. Values below 1.0 indicate undersensitivity; values above 1.0 indicate oversensitivity to changes in difficulty. Table 2 shows the means and ranges of this sensitivity ratio for the three groups of subjects. All subjects were undersensitive; however, here (at last!) we find a clear superiority of experts over the other two groups. Only two of the untrained subjects and three of the trained subjects were more sensitive

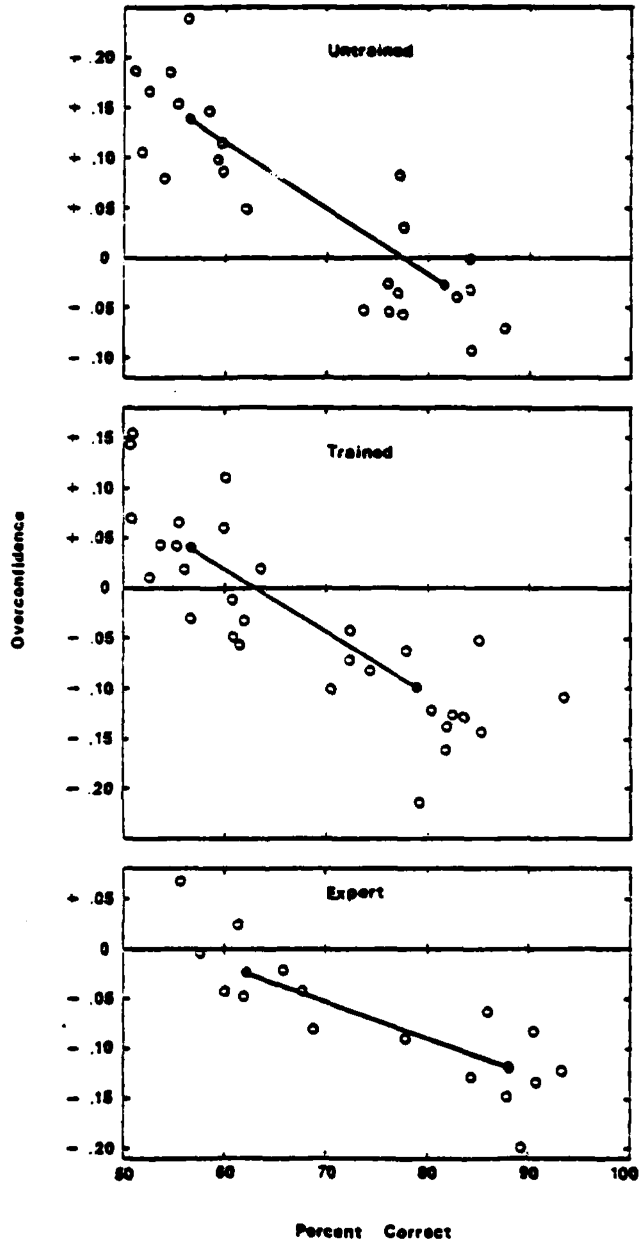


FIGURE 3. Overconfidence plotted against percentage correct for each subject on the hard (circles) and easy (triangles) tests. The solid symbols are group means.

to changes in difficulty than was the least sensitive expert. The trained subjects were not significantly better than the untrained group on this measure.[3]

Discussion

At first glance, our probability experts appeared no better calibrated than the untrained group; both were inferior to the trained group. Most of the untrained subjects were overconfident, whereas the experts were underconfident.

We will never know the calibration of our experts before they started studying the calibration of others. However, some archival data seem relevant. The first calibration data we ever collected used 19 employees of the Oregon Research Institute as subjects. Those subjects were similar to our present experts in that they were psychologists studying human judgment and their equally knowledgeable secretaries and research assistants. But they knew nothing about calibration. Their responses showed the now-familiar severe overconfidence (Lichtenstein et al., 1977, Figure 6). The underconfidence of the present experts seems to represent a change in behavior prompted by research findings. Apparently, they were determined not to be overconfident and, in their zeal, they over-corrected, becoming underconfident over a wide range of difficulty levels.

If this interpretation is correct, these results are moderately encouraging. People <u>can</u> learn from the experience of others. Readers of this article, having learned both of the general tendency to be overconfident and of the possibility of overcorrection, should therefore be able to produce well-calibrated responses.[4]

Table 2

Sensitivity Ratio

	Group		
	Untrained	Trained	Experts
Mean	.37	.30	.60
Range	.04 to .68	.02 to .65	.45 to .73

A more detailed examination of the data indicated that the labeling of one group of assessors as "better calibrated" than another is rendered moot by the systematic relationship between difficulty and overconfidence. This relationhsip is apparently mediated by an insensitivity to changes in difficulty. The experts were superior to both other groups in being more sensitive to changes in difficulty of the items, but the hard/easy effect was still substantial in the expert group.

The hard/easy effect can be viewed as a regression effect: If the correlation between the ease of item (defined by some objective criterion) and proportion correct is greater than the correlation between ease of item and mean response, as shown in Figure 4, then for easy items, one would expect to observe underconfidence ($P_e > \Psi_e$), whereas for the hard items, one would expect overconfidence ($\Psi_h > P_h$).

Some introspection suggests how hard it would be to be fully sensitive to test difficulty. Whether defined in terms of the proportion of people getting an item correct (as in Lichtenstein & Fischhoff, 1977) or by reference to some measure related to an item's content (as done here), item difficulty does not appear to be a piece of information above and beyond one's general feelings of uncertainty about which answer is correct. Intuitively, easy items are just those to which one is inclined to state a high probability, while hard items are those about which one is quite unsure. Sensitivity to difficulty might require a counter-intuitive, two-stage process, involving the assessment of both personal uncertainty and item difficulty ("I'm pretty sure I know the answer, but this seems like a hard item, so I'd better lower my confidence") before arriving at a probability assessment. Indeed,

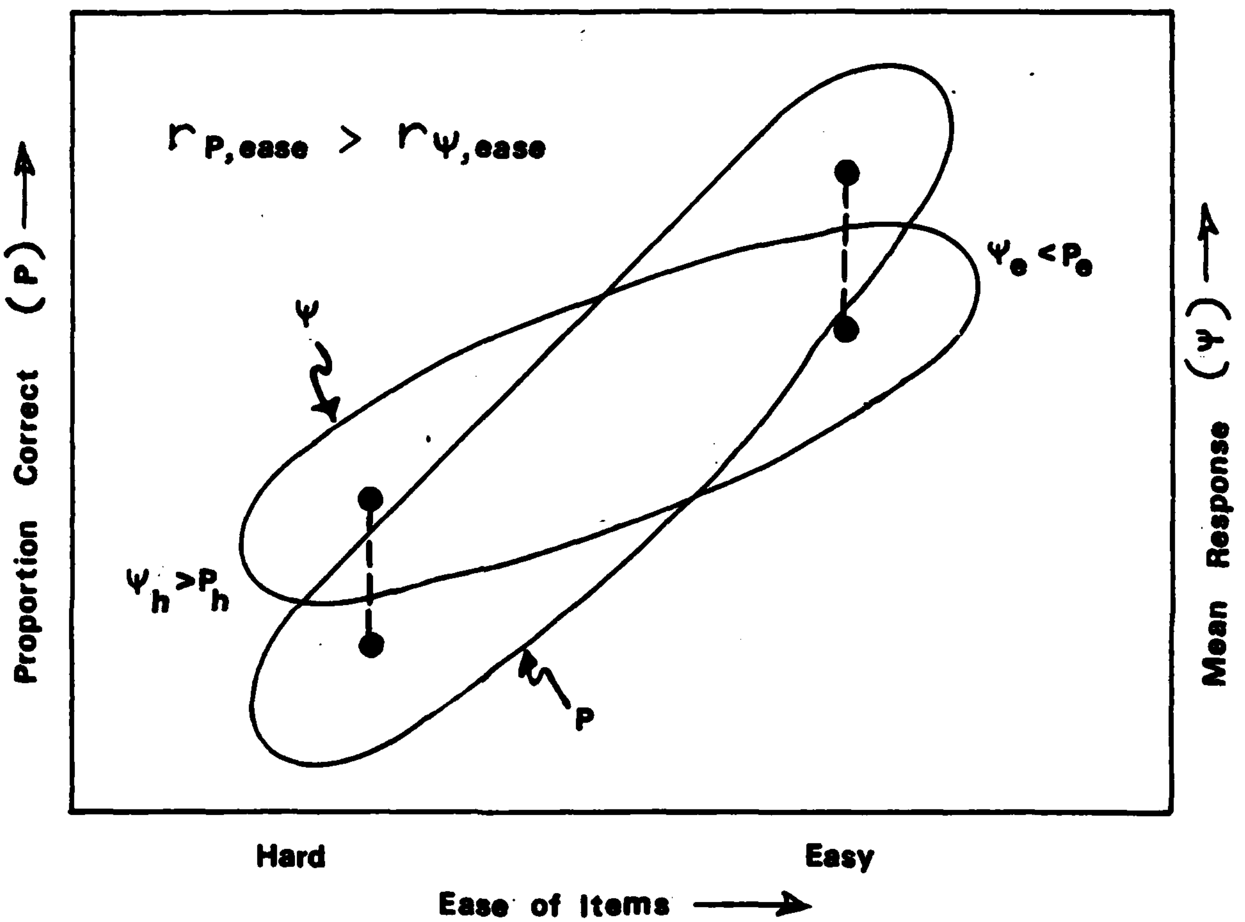


Figure 4. A schematic representation of how the hard/easy effect would result from a higher correlation between difficulty and percentage correct than between difficulty and mean response.

although both the present authors were above average (even among the experts) in sensitivity to difficulty, we did not consciously use such a two-stage process, and do not know how we did as well as we did.

Reference Note

1. Cambridge, R. M. & Shreckengost, R. C. Are you sure? The subjective probability assessment test. Unpublished manuscript, Langley, Virginia: Office of Training, Central Intelligence Agency, 1978.

References

Clarke, F. R. Confidence ratings, second-choice responses, and confusion matrices in intelligibility tests. Journal of the Acoustical Society of America, 1960, 32, 35-46.

Fischhoff, B. & Slovic, P. A little learning. . .: Confidence in multicue judgment. In R. Nickerson (Ed.), Attention and Performance, VIII. Hillsdale, N. J.: Erlbaum, in press.

Fischhoff, B., Slovic, P. & Lichtenstein, S. Knowing with certainty: The appropriateness of extreme confidence. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 552-564.

Jungermann, H. Aiding individuals to make difficult decisions. Acta Psychologica, 1980, 45.

Koriat, A., Lichtenstein, S. & Fischhoff, B. Reasons for confidence. Journal of Experimental Psychology: Human Learning and Memory, 1980, 6, 107-118.

Lichtenstein, S. & Fischhoff, B. Do those who know more also know more about how much they know? The calibration of probability judgments. Organizational Behavior and Human Performance, 1977, 20, 159-183.

Lichtenstein, S. & Fischhoff, B. Training for calibration. Organizational Behavior and Human Performance, 1980.

Lichtenstein, S., Fischhoff, B. & Phillips, L. D. Calibration of probabilities: The state of the art. In H. Jungermann & G. de Zeeuw (Eds.), Decision making and change in human affairs. Dordrecht, Holland: D. Reidel, 1977.

Murphy, A. H. Scalar and vector partitions of the probability score (Pt.1). Two-state situation. Journal of Applied Meteorology, 1972, 11, 273-282.

Murphy, A. H. A new vector partition of the probability score. Journal of Applied Meteorology, 1973, 12, 595-600.

Nickerson, R. S. & McGoldrick, C. C. Jr. Confidence ratings and level of performance on a judgmental task. Perceptual and Motor Skills, 1965, 20, 311-316.

Pitz, G. F. Subjective probability distributions for imperfectly known quantities. In L. W. Gregg (Ed.), Knowledge and cognition. New York: Wiley, 1974.

U. S. Nuclear Regulatory Commission. Reactor safety study: An assessment of accident risks in U. S. commercial nuclear power plants. WASH 1400 (NUREG-75/014). Washington, D. C.: The Commission, 1975.

Wright, G. N., Phillips, L. D., Whalley, P. C., Choo, G. T. G., Ng, K. O., Tan, I., & Wisudha, A. Cultural differences in probabilistic thinking. Journal of Cross-cultural Psychology, 1978, 9, 285-299.

Footnotes

We acknowledge with thanks our discussions with Paul Slovic and Daniel Kahneman. This research was supported in part by the Advanced Research Projects Agency of the Department of Defense, monitored by the Office of Naval Research under Contract N00014-79-C-0029 (ARPA Order No. 3668) to Perceptronics, Inc., and in part from the Office of Naval Research under Contract N00014-80-C-0150 to Perceptronics, Inc. Requests for reprints may be addressed to Sarah Lichtenstein, Decision Research, 1201 Oak Street, Eugene, Oregon 97401.

1. Our deep thanks to Barbara Combs, Dennis Fryback, Barbara Goodman, Don MacGregor, Gordon Pitz, and David Seaver for serving as our expert subjects.

2. Because this measure is artifactually increased by the infrequent use of two-digit probabilities (e.g., .95), the data were grouped (.5-.59, .6-.69, ... , .9-.99, 1.0) for calculating the calibration index.

3. One-way ANOVA: $F = 10.4$; $p = .0003$. Trained vs. untrained, $t = 1.20$, $p > .2$; trained vs. expert, $t = 3.57$, $p = .001$; untrained vs. expert, $t = 4.44$, $p = .0001$.

4. We are willing to analyze the results for the first 40 readers who write us to accept this challenge. You will have the advantage of knowing the approximate difficulty of the test.

DISTRIBUTION LIST

OSD

CDR Paul R. Chatelier
Office of the Deputy Under
Secretary of Defense
OUSDRE (E&LS)
Pentagon, Rm. 3D129
Washington, D. C. 20301

Department of the Navy

Director
Engineering Psychology Programs
Code 455
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217 (5 copies)

Director
Tactical Development &
Evaluation Support
Code 230
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

Director
Naval Analysis Programs
Code 431
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

Director
Operations Research Programs
Code 434
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

Director
Statistics and Probability Program
Code 436
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

Director
Information Systems Program
Code 437
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

Department of the Navy

Special Assistant for Marine
Corps Matters
Code 100M
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

Commanding cfficer
ONR Eastern/Central Regional Office
ATTN: Dr. J. Lester
Bldg. 114, Section D
666 Summer Street
Boston, Massachusetts 02210

Commanding Officer
ONR Branch Office
ATTN: Dr. C. Davis
536 South Clark Street
Chicago, Illinois 60605

Commanding Officer
ONR Western Regional Office
ATTN: Dr. E. Gloye
1030 East Green Street
Pasadena, California 91106

Office of Naval Research
Scientific Laison Group
American Embassy, Room A-407
APO San Francisco, California 96503

Director
Naval Research Laboratory
Technical Information Division
Code 2627
Washington, D. C. 20375 (6 copies)

Dr. Bruce Wald
Communications Sciences Division
Code 7500
Naval Research Laboratory
Washington, D. C. 20375

Dr. Robert G. Smith
Office of the Chief of Naval
Operations, OP987H
Personnel Logistics Palns
Washington, D. C. 20350

Naval Training Equipment Center
ATTN: Technical Library
Orlando, Florida 32813

Department of the Navy

Human Factors Department
Code N215
Naval Training Equipment Center
Orlando, Florida 32813

Dr. Alfred F. Smode
Training Analysis and
Evaluation Group
Naval Training Equipment Center
Code N-OOT
Orlando, Florida 32813

Dr. Arthur Bachrach
Behavioral Sciences Department
Naval Medical Research Institute
Bethesda, Maryland 20014

Dr. George Moeller
Human Factors Engineering Branch
Submarine Medical Research Lab
Naval Submarine Base
Groton, Connecticut 06340

Commanding Officer
Naval Health Research Center
San Diego, California 92152

Dr. James McGrath, Code 311
Navy Personnel Research and
Development Center
San Diego, California 92152

Navy Personnel Research and
Development Center
Management Support Department
Code 210
San Diego, California 92152

CDR P. M. Curran
Code 604
Human Factors Engineering Division
Naval Air Development Center
Warminster, Pennsylvania 18974

Mr. Ronald A. Erickson
Human Factors Branch
Code 3194
Naval Weapons Center
China Lake, California 93555

Human Factors Engineering Branch
Code 1226
Pacific Missile Test Center
Point Mugu, California 93042

Department of the Navy

Dean of the Academic Departments
U.S. Naval Academy
Annapolis, Maryland 21402

Dr. Gary Poock
Operations Research Department
Naval Postgraduate School
Monterey, California 93940

Dean of Research Administration
Naval Postgraduate School
Monterey, California 93940

Mr. Warren Lewis
Human Engineering Branch
Code 8231
Naval Ocean Systems Center
San Diego, California 92152

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps
Code RD-1
Washington, D. C. 20380

HQS, U.S. Marine Corps
ATTN: CCA40 (MAJOR Pennell)
Washington, D. C. 20380

Mr. Phillip Andrews
Naval Sea Systems Command
NAVSEA 0341
Washington, D. C. 20362

Commander
Naval Electronics Systems Command
Human Factors Engineering Branch
Code 4701
Washington, D. C. 20360

LCDR W. Moroney
Code 55MP
Naval Postgraduate School
Monterey, California 93940

Mr. Merlin Malehorn
Office of the Chief of Naval
Operations (OP 102)
Washington, D. C. 20350

Department of the Army

Mr. J. Barber
HQS, Department of the Army
DAPE-MBR
Washington, D. C. 20310

Department of the Army

Dr. Joseph Zeidner
Technical Director
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, Virginia 22333

Director, Organizations and
Systems Research Laboratory
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, Virginia 22333

Technical Director
U.S. Army Human Engineering Labs
Aberdeen Prooving Ground, Md. 21005

U.S. Army Aeromedical Research Lab
ATTN: CPT Gerald P. Krueger
Ft. Rucker, Alabama 36362

ARI Field Unit-USAREUR
ATTN: Library
C/O ODCSPER
HQ USAREUR & 7th Army
APO New York 09403

Department of the Air Force

U.S. Air Force Office of
Scientific Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, D. C. 20332

Dr. Donald A. Topmiller
Chief, Systems Engineering Branch
Human Engineering Division
USAF AMRL/HES
Wright-Patterson AFB, Ohio 45433

Air University Library
Maxwell Air Force Base, Al 36112

Dr. Gordon Eckstrand
AFHRL/ASM
Wright-Patterson AFB, Ohio 45433

Foreign Addresses

North East London Polytechnic
The Charles Myers Library
Livingstone Road
Stratford
London E15 2LJ
ENGLAND

Foreign Addresses

Prof. Dr. Carl Graf Hoyos
Institute for Psychology
Technical University
8000 Munich
Arcisstr 21
FEDERAL REPUBLIC OF GERMANY

Dr. Kenneth Gardner
Applied Psychology Unit
Admiralty Marine Technology
Establishment
Teddington, Middlesex TW11 OLN
ENGLAND

Director, Human Factors Wing
Defense and Civil Institute of
Environmental Medicine
P. O. Box 2000
Downsview, Ontario M3M 3B9
CANADA

Dr. A. D. Baddeley
Director, Applied Psychology Unit
Medical Research Council
15 Chaucer Road
Cambridge, CB2 2EF
ENGLAND

Other Governmental Agencies

Defense Documentation Center
Cameron Station, Bldg. 5
Alexandria, Virginia 22314 (12 copies)

Dr. Craig Fields
Director, Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, Virginia 22209

Dr. Judith Daly
Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, Virginia 22209

Professor Douglas E. Hunter
Defense Intelligence School
Washington, D. C. 20374

Other Organizations

Dr. Robert R. Mackie
Human Factors Research
5775 Dawson Avenue
Goleta, California 93017

Dr. Gary McClelland
Institute of Behavioral Sciences
University of Colorado
Boulder, Colorado 80309

Human Resources Research Office
300 N. Washington Street
Alexandria, Virginia 22314

Dr. Miley Merkhofer
Stanford Research Institute
Decision Analysis Group
Menlo Park, California 94025

Dr. Jesse Orlansky
Institute for Defense Analyses
400 Army-Navy Drive
Arlington, Virginia 22202

Prof. Judea Pearl
Engineering Systems Department
University of California
405 Hilgard Ave.
Los Angeles, California 90024

Prof. Howard Raiffa
Graduate School of Business Administration
Harvard University
Soldiers Field Road
Boston, Massachusetts 02163

Dr. T. B. Sheridan
Dept. of Mechanical Engineering
MIT
Cambridge, Massachusetts 02139

Dr. Arthur I. Siegel
Applied Psychological Services
404 East Lancaster Street
Wayne, Pennsylvania 19087

Dr. Amos Tversky
Department of Psychology
Stanford University
Stanford, California 94305

Other Organizations

Dr. Gershon Weltman
Perceptronics
6271 Variel Avenue
Woodland Hills, California 91364

Dr. Meredith P. Crawford
American Psychological Association
Office of Educational Affairs
1200 - 17th Street, N. W.
Washington, D. C. 20036

Dr. Ward Edwards
Director, Social Science Research Institute
University of Southern California
Los Angeles, California 90007

Dr. Charles Gettys
Department of Psychology
University of Oklahoma
455 West Lindsey
Norman, Oklahoma 73069

Dr. Kenneth Hammond
Institute of Behavioral Science
University of Colorado
Room 201
Boulder, Colorado 80309

Dr. Ronald Howard
Department of Engineering-Economic Systems
Stanford University
Stanford, California 94305

Dr. William Howell
Department of Psychology
Rice University
Houston, Texas 77001

Journal Supplement Abstract Service
American Psychological Association
1200 - 17th Street, N. W.
Washington, D. C. 20036 (3 copies)

Mr. Richard J. Heuer, Jr.
27585 Via Sereno
Carmel, California 93923

Mr. Tim Gilbert
The MITRE Corporation
1820 Dolly Madison Blvd.
McLean, Virginia 22102

Other Organizations

Dr. John Payne
Duke University
Graduate School of Business
Administration
Durham, North Carolina 27706

Dr. Andrew P. Sage
University of Virginia
School of Engineering and
Applied Science
Charlottesville, Virginia 22901